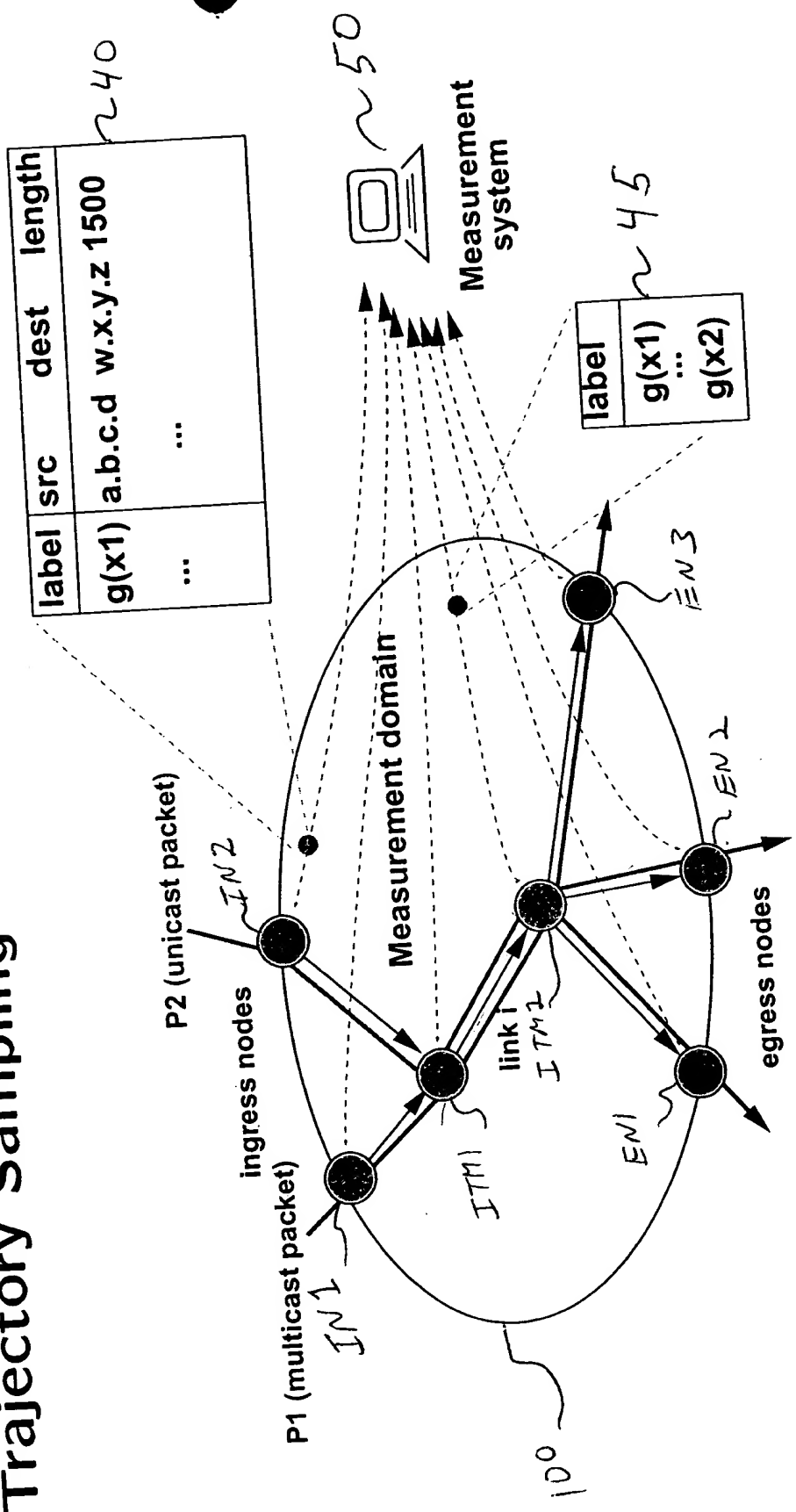


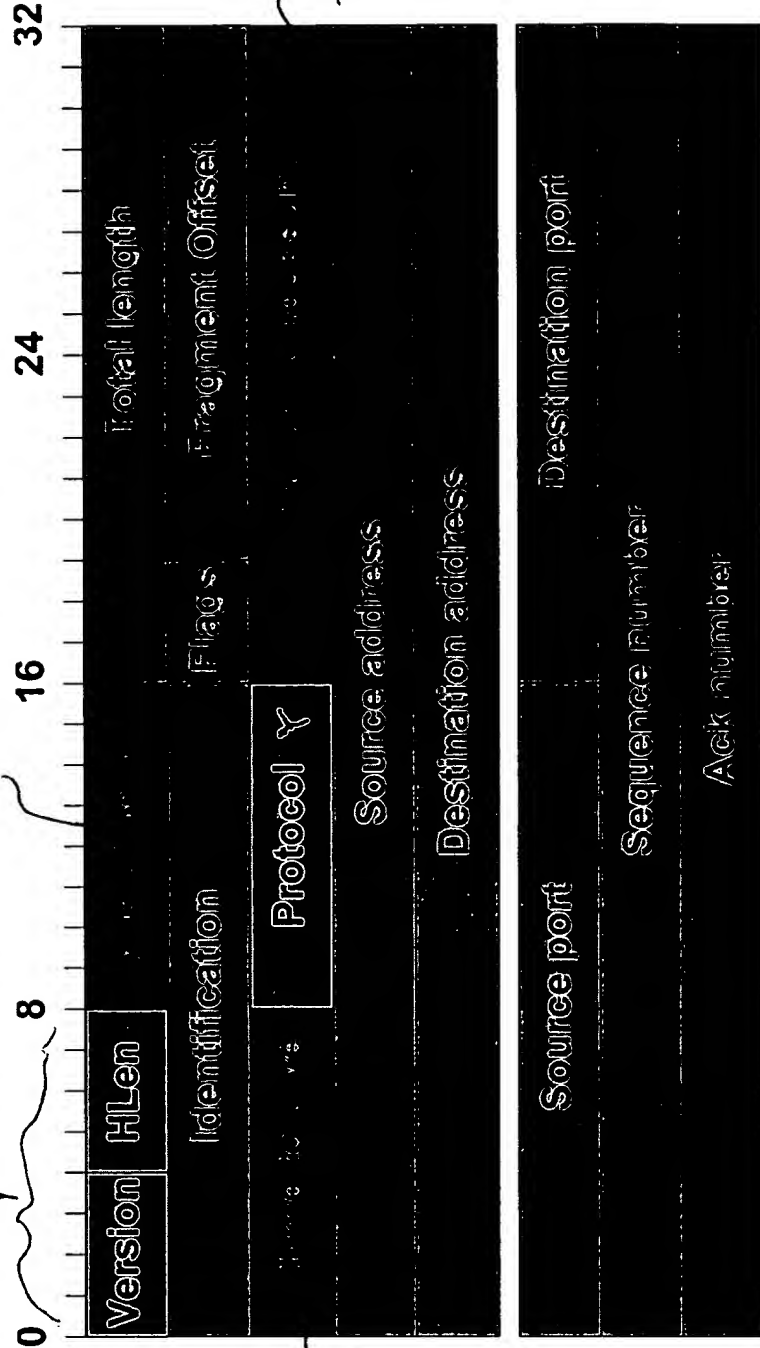
60464 RE 960

# Trajectory Sampling



- Collect fields of interest only once (ingress)
- Multicast requires no special treatment

# Fields Included in Hashes <sup>R</sup>



IP header

TCP/UDP

Low-entropy  $R$

Low-entropy  $\gamma$

High-entropy  $G$

09688335-101600

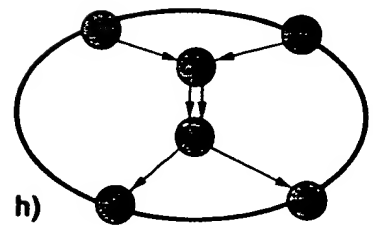
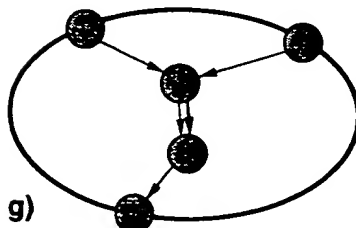
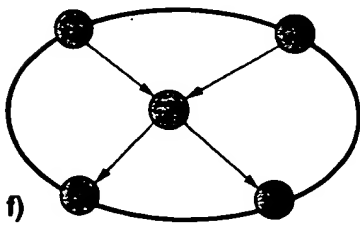
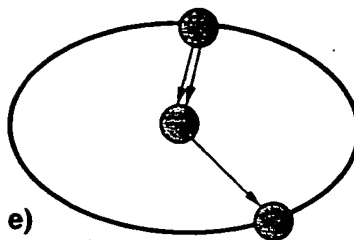
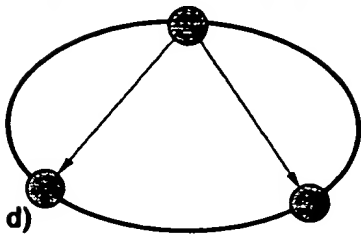
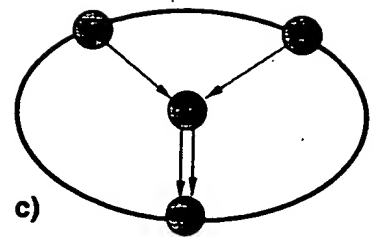
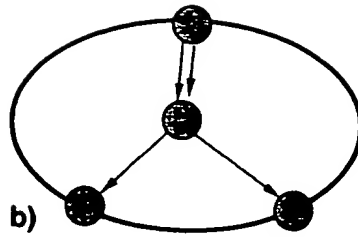
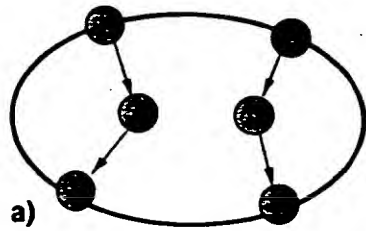
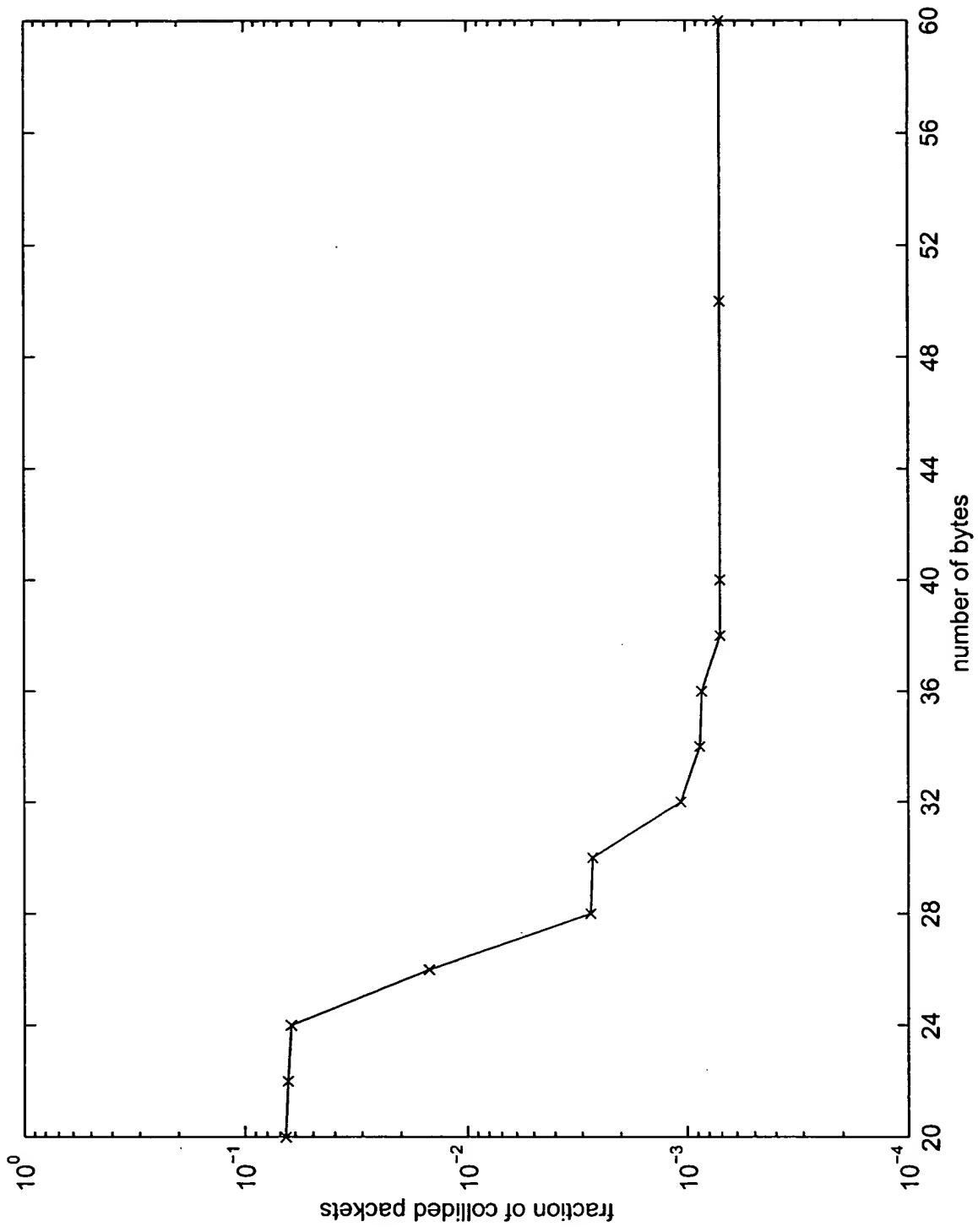


Figure 3

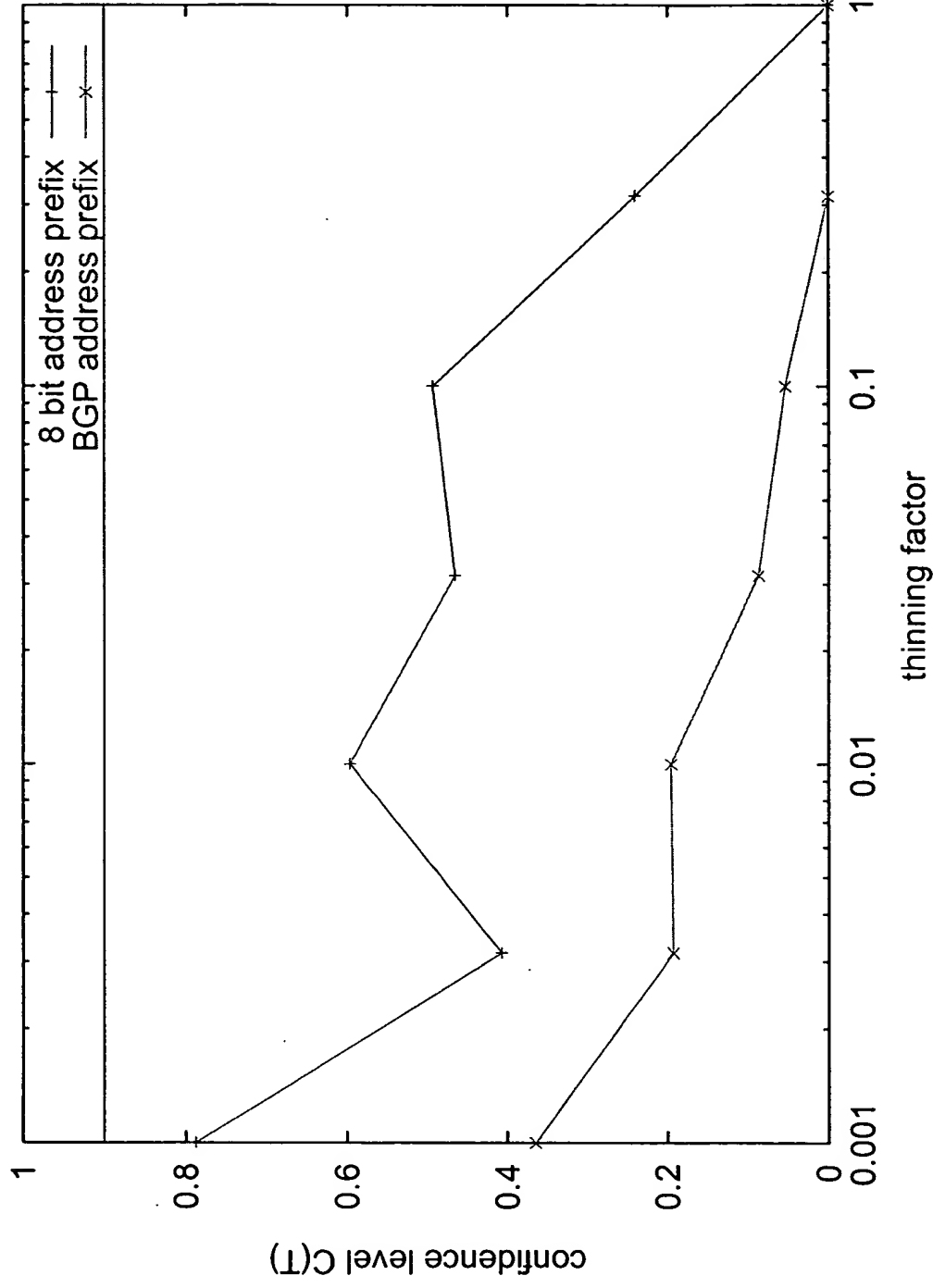
00916-00000000

## Collisions: Identical Packets are Rare



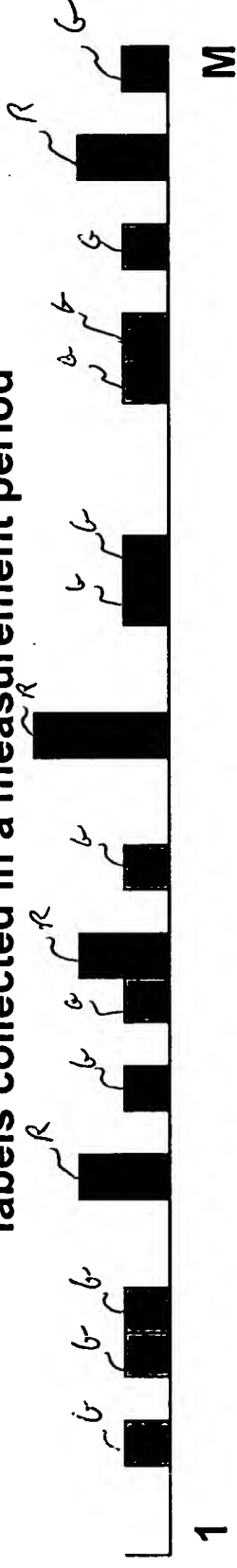
# $\chi^2$ -Test for Independence of Sampling Decision & Addresses

- If  $C(T) < 1 - \text{significance level} \Rightarrow \text{accept hypothesis}$



# Optimal Sampling

labels collected in a measurement period



- Fix amount of measurement traffic  $c$  per period
- Tradeoff: collisions vs. label size
- Problem:
  - $n$ : number of samples in sampling period
  - $M$ : alphabet size,  $m = \log_2 M$  [bits/label]
  - $n \cdot m$ : total amount of measurement traffic [bits]
  - Goal: maximize number of *unique* labels  
subject to  $n \cdot m \leq c$ .
- Optimal alphabet size:  $M^* = c \log(2)$
- Optimal number of samples:  $n^* = \frac{M^*}{\log(M^*)}$

Example:  $c = 10^6$  bit  $\Rightarrow m^* = 19.4$  bit/label

$$n^* = 5.15 \cdot 10^4 \text{ samples}$$

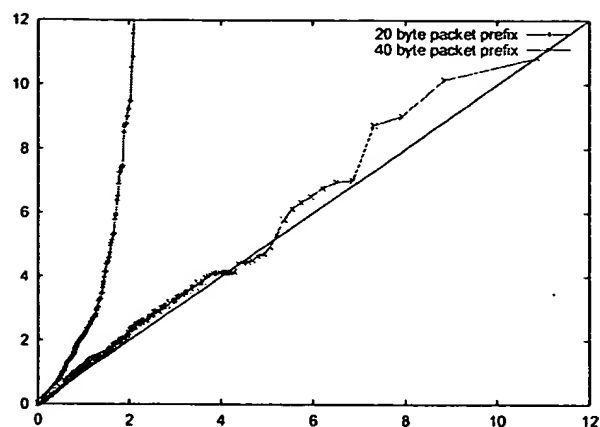


Figure 6: HASH-SAMPLED ADDRESS BITS DISTRIBUTIONS. Quantile-quantile plot of address bit chi-square values vs. chi-squared distribution with 1 degree of freedom; for various traces, primes  $A$ , thinning factors  $r/A$ ; see text. Close agreement for 40 byte packet prefixes; marked disagreement for 20 byte packet prefixes (i.e. no payload included for sampling hash)

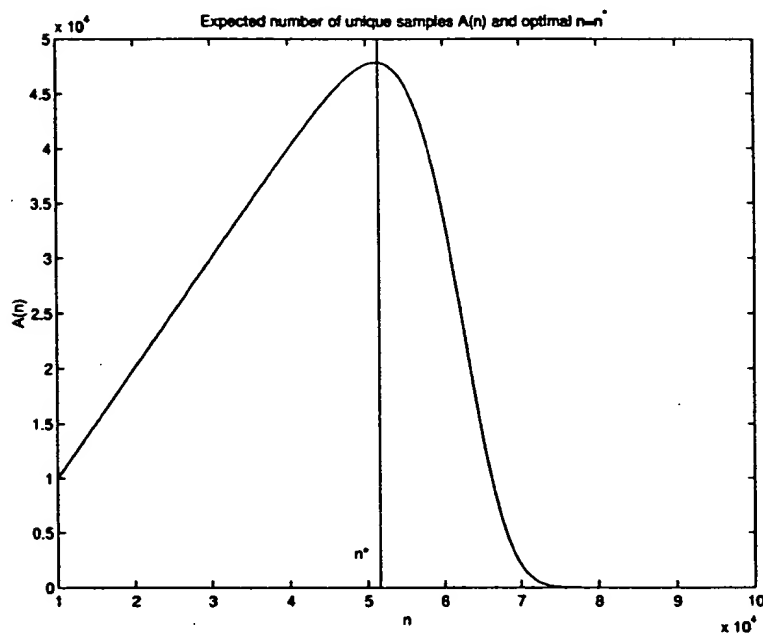
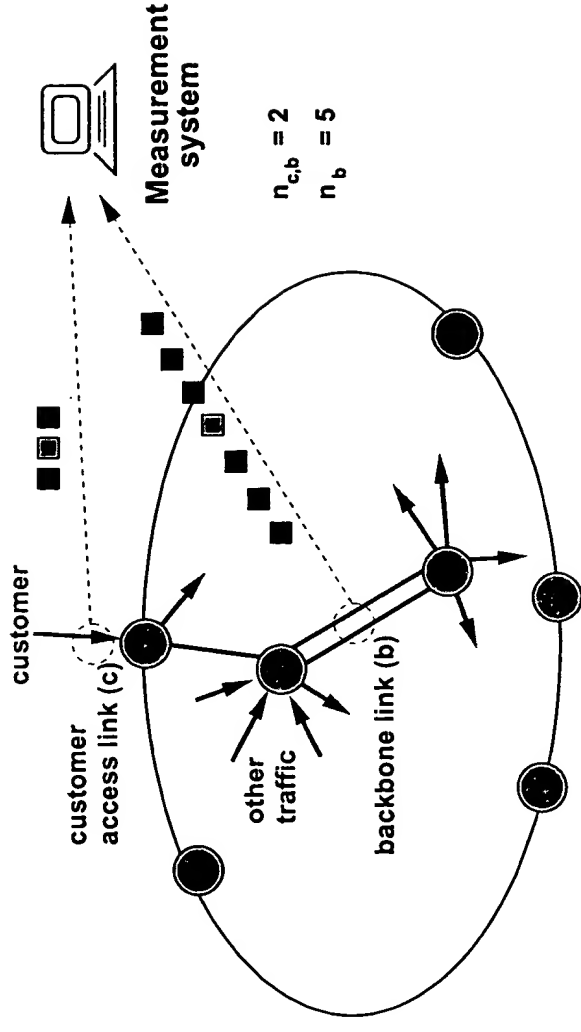


Figure 7: The expected number of unique samples  $A(n)$  as a function of  $n$ , for  $c = 10^6$  bit. The optimal number of samples  $n^*$  is approximately  $5.15 \cdot 10^4$ , with  $m^* = 19.4$  bit per label. The collision probability  $p_{coll}$  is approximately 0.072, i.e., 7.2% of the samples transmitted to the collection system have to be discarded.

# Inference Experiment

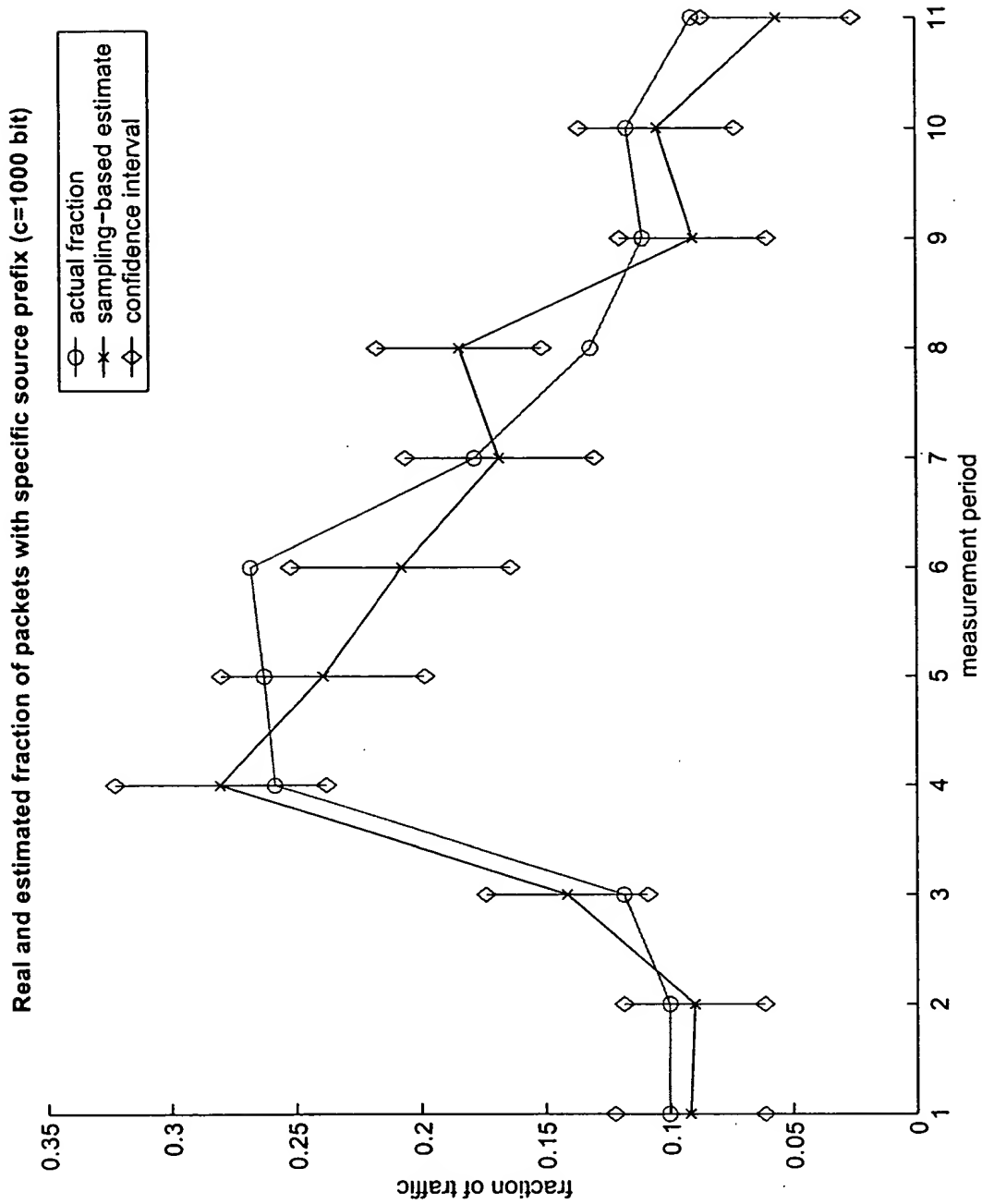
- Experiment: inference from trajectory samples
  - Estimate fraction of traffic from customer
  - Customer traffic: small source address subset



- Fraction of customer traffic on backbone:  $\mu$   
 Estimator:  $\hat{\mu} = n_{c,b}/n_b$ 
  - $n_{c,b}$ : # unique labels common on both links
  - $n_b$ : # unique labels on backbone link
- Ingress link and source address correlated



# Estimated Customer Traffic ( $c = 10^3$ [bits/epoch])



# Estimated Customer Traffic ( $c = 10^4$ [bits/epoch])

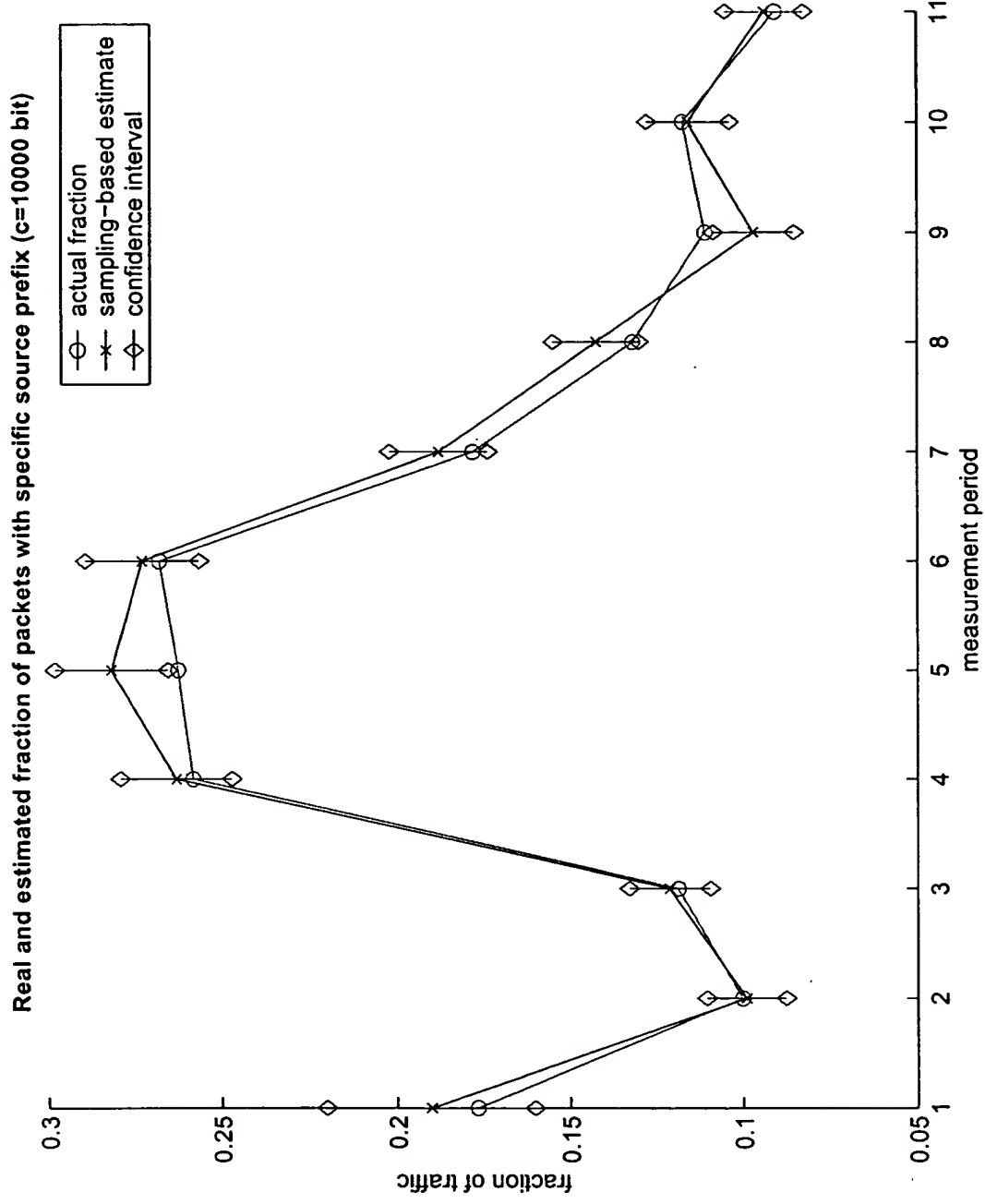


FIGURE 11

# Sampling Device Implementation

~ 1100

